

Future of CF conventions and standard names

Jonathan Gregory

CGAM, Department of Meteorology, University of Reading, UK
Met Office Hadley Centre, UK

CF was developed by Brian Eaton, Jonathan Gregory, Bob Drach,
Karl Taylor and Steve Hankin.

Views expressed herein are my own!

Goals

Locate data in space–time and as a function of other independent variables, to facilitate processing and graphics

Identify data sufficiently to enable users of data from different sources to decide what is comparable, and to distinguish variables in archives

Intended for:

use with climate and forecast data

atmosphere, surface and ocean

model-generated data and comparable observational datasets

General principles

Data should be self-describing—no external tables needed to interpret it

Conventions have been developed only for things we know we need

Avoid being too onerous for data-writers and data-readers

Metadata readable by humans as well as easily parsed by programs

Minimise redundancy and possibilities for silly mistakes e.g. *when CF prescribes a set of possible values for an attribute, they are not case-sensitive; this saves mistakes by data-writers but makes slightly more work for data-readers*

Growing use

CF is being used by an increasing number of groups

- because they like it and it suits their purposes and/or
- even when reluctantly, because it is becoming a de-facto standard.

Google has 13400 hits for “CF netCDF standard”. (*CF home page is number 1.*)

Users include BADC, CloudNET radar programme, GLIMMER community ice sheet model project, GODAE (ocean data assimilation), HOAPS (Hamburg satellite-delivered freshwater flux climatology), JISAO climate data archive, Marine Metadata Interoperability project (MMI), US Naval Oceanographic Office, PMIP II, MERSEA (European marine environment) EU project, IPCC (through PCMDI) and hence all AOGCMs, PRISM.

Wider use ⇒

- More requests for development.
- Need for faster response *to avoid fragmentation of the community—if we do not deal with issues fast enough, groups of users will make private modifications to the standard for their own purposes, and this will undermine its role as a general standard.*
- Care is necessary in design and responsibility in decision-making *since many people will be affected.*

All of these in turn ⇒ **more effort is required**, of two kinds: people specifically employed to work on CF, and engagement by others who care about it. *The former need is now desperate, in my opinion, to ensure the continued usefulness of CF by building on what we have already done. It is worth noting that no-one has ever had CF as part of their job officially. It has always been a spare-time activity, and not much time is spare. For instance, I am spending ~ 5% of my time on it.*

Secretary of the CF standard

Someone should be employed to update and maintain the standards document. *The minimum requirement is to keep it up to date with our decisions as they are made. The compliance document, which defines what is a violation of the standard, should also be maintained. Brian Eaton has been carrying out this role for some time but does not feel able to continue with it.*

Changes which have been agreed but not yet implemented *in the standard as a consequence of no-one having time to do it.*

- Extension to `cell_methods` and definition of `where` in standard names.
- Standard name parameters e.g. *reference pressure for potential temperature, geoid name for quantities relative to the geoid, reference power level for quantities in decibels.*
- Relation of `formula_terms` and `bounds`.
- Relation of forecast and validity time.

The secretary could also formulate proposals as necessary to encapsulate ideas which have been discussed and hence move the discussion forward.

Developing the conventions

The procedure adopted should be as simple as possible, but no simpler.

Media for proposals:

- Discuss and propose changes on CF general email list (as now).
- Maintain a web archive of proposed amendments (thread, branch, version).

I'm not convinced we need special discussion groups for particular issues *as discussions often branch and interact, and threads in the email list can be followed from the subject lines.*

Agreeing changes

After one month of no new proposal within the thread, acceptance or decision by CF “guardians”. *Usually there is no need for a decision between alternatives, since the discussion generally leads to a consensus, with withdrawal of all but one of the proposals. It has rarely been necessary to take a vote.*

Guardians should

- have a veto.
- accept responsibility for the integrity of CF.
- be identified on the CF website.
- represent a range of communities.

In proposing an idea like this, Bryan Lawrence called these people a “security council”.

New major version of standard if a section or appendix added (or deleted), minor version otherwise. *By this criterion we are currently heading for CF-1.1.*

Publish change on new CF news/announcement email list.

Names of proposers of agreed changes should be added to CF author list.

Issues for development of the standard

- **Grid mappings and GIS** *i.e.* relationship between map (x, y) coordinates and geographical lat–lon coordinates, identified in the [grid_mapping](#) variable *i.e.* *Appendix F*. Not really related to “grid”. More general than “map projection” *because it includes rotated pole, for instance.*
- **Wholesale adoption of other standards.** Problems: other standards may not be so precise as ours, may not fit our standard, need to keep in step. *E.g.* GCMD region names *could not be adopted directly because they didn't have quite the selection of contents or format that we needed.*
- **Grid structure** *i.e.* how indices of data variable relate to geographical location and relative location of points. *E.g.* conformal cubic, tripolar. Different from grid mappings. Provide a new ancillary variable like [grid_mapping](#)?
- **Staggered grids.** Do we need to record the relation between various grids used in a file? Problems: redundancy (hence potential inconsistency), variables not self-contained (what to do when combining files?)

Backward compatibility

CF is backward-compatible with COARDS so that

- applications which understand CF can also process COARDS datasets.
- CF datasets will not break applications based on COARDS.

Hence:

COARDS is a subset of CF

where COARDS is adequate, CF does not provide an alternative
extensions to COARDS are all optional and provide new functionality
alternatives to COARDS cannot be depended on e.g. to identify axes
it is hard to say what “CF-compliance” means beyond COARDS support

Should we be more demanding?

Secretary of CF standard names (Kurremkarmerruk, *the Master Namer*)

Standard names are an optional part of CF. CF should not be judged by standard names alone! Standard names have wider application than the CF standard. *I am painfully aware that the standard names and guidelines are largely my own work and ideas, with which others may disagree.*

Deciding standard names requires a lot of time and intellectual effort, because it is necessary to understand what they mean.

Someone should be employed to check and implement requests for new standard names, and to write explanatory comments for them. This person will need to seek guidance from experts in relevant areas. *Such scientific input is needed in particular because many requests from standard names come from people who are trying to put data into CF for the first time but may not personally be able to describe its contents scientifically with the precision required by standard names.*

Questions to be considered for new standard names:

- What does it mean? E.g. *I had to educate myself about “dry static energy” and “spherical irradiance” in order to discuss these names with their proposers.*
- Is it really what is intended? E.g. *a recent request for a name of “heat content” was clarified as being the integral of sea water temperature wrt depth.*
- Is this the same as one we already have? E.g. *“sea ice production” was clarified as being [tendency_of_sea_ice_thickness_due_to_thermodynamics](#), already in the table.*
- How should it most clearly be named, using familiar terms if possible? *We chose to use [flux](#) to mean “flux density” and [transport](#) to mean “flux” (in physics terms) following common use.*
- How should it best be named in keeping with other quantities?

Principles of standard names

- We assign names when they are needed and not before. *The guidelines try to anticipate future needs, but are not binding until used.*
- We will give a name to every distinct quantity which users wish to identify, acknowledging that different applications draw different distinctions. *This is like different ways of naming colours. In one application we might group colours as black, white, red, green and blue; more subtly we would add grey, brown, purple, light and dark variants, etc.; paint manufacturers distinguish a huge number of hues!*
- No built-in categories or aggregations. *Any categorisation will have drawbacks. For example, is “surface net downward radiation” a radiation quantity or a surface quantity? The current CF standard name categories are overlapping and computed on the fly; the standard name list itself is “flat”.*
- Different physical dimensions \Rightarrow different standard name, but physically equivalent units are acceptable ($\text{cm} \equiv \text{m}$, $\text{dbar} \equiv \text{Pa}$).
- Ambiguity avoided e.g. no “surface downward longwave flux” *because it would be unclear whether it meant [surface_downwelling_longwave_flux](#) or [surface_net_downward_longwave_flux](#), so these names instead are the ones we have in the table.*
- Sign and direction specified in standard name e.g. “downward” or “upward”—*it is vital that this information is provided, and it cannot be omitted if we make it part of the name.*
- Spatiotemporal or other ancillary data not included in standard name e.g. *1.5 m, temporal min/max/mean, Atlantic ocean, radiation wavelength ranges—however, surfaces are identified by the standard name.*
- Context specified and jargon avoided because CF has a wide scope e.g. *we distinguish [eastward_wind](#) and [eastward_sea_water_velocity](#), which might both be called u , and we have renamed [omega](#) as [langrangian_tendency_of_air_pressure](#).*
- Systematic and consistent naming, using precedent and guidelines specifying construction rules e.g.

[surface] [component] standard_name [at surface] [in medium] [due to process] [where type] [assuming condition]

also transformations, special phrases and generic names.

Defining files

Standard names are currently defined as follow:

`standard_name_list` specifies names and canonical units

`standard_name_aliases` gives equivalences of superseded names to current names (to correct mistakes and make changes)

`standard_name_comments` provides explanatory text to be selected by matching regular expressions against standard names

From these and some other files we generate:

`standard_name.html`: web page with formatted table of names and aliases and various links, help text generated on request by a cgi script.

`standard_name.xml`: file with an entry for each alias and name, and the help text generated for each name.

Agreeing new standard names

CF users should request new standard names by either

- Generating them automatically from a web form (to be set up), which will automatically post them to email list and change-control system. *They could not be accepted automatically because some nonsensical names might be generated, or the quantity might already have a different name in the table.*
- Sending them to the secretary of CF standard names, who will post them after discussing with proposer and seeking expert guidance if needed. *This is generally what has happened in practice.*

If no objection is made within two weeks, accept proposal and update the standard name table etc.

Publish new names on new CF news/announcement email list.

Proposers of new standard names which can't be generated automatically should be listed as contributors.

Projects and interest groups should be encouraged to propose new groups of names. We could delegate responsibility for proposing names in their areas to such groups. But I believe **there must always be central ratification and a single table**—otherwise there is no point in standard names. *If for instance we delegated authority to “ocean” and “atmosphere” groups, it is likely they would introduce ambiguous names e.g. “northward heat transport” might occur in both lists, and they might separately give names to some quantities of common interest e.g. surface windstress. These kinds of difficulty would mean names were no longer unique and precise.*

Issues for development of the standard name table

- Equivalences with other code tables. Who is responsible? One-to-one correspondence may not be possible. On-line tables or multiple codes in files?
- Observed quantities which depend on the instrument or measurement technique, not just on properties of the observed world e.g. cloud radar signal-to-noise ratio, ADCP acoustic intensity, any “raw” data.
- Extensions needed to `udunits.dat` e.g. decibel sverdrup psu. Maintain a CF version?
- Use of parameters or dimensions for “multivalued” parts of standard names e.g. chemical species, plant functional types. *BODC parameters have to address such complexity e.g. in the biological domain.*
- Complicated derived quantities e.g. “growing season length” defined as period between when $T_{\text{day}} > 5^{\circ}\text{C}$ for > 5 d and $T_{\text{day}} < 5^{\circ}\text{C}$ for > 5 d.

Other supporting activities

- Maintain the CF web page.
- Keep the CF checker up-to-date with the standard. Extend it to provide a description of the data, using all CF metadata. *In that case it would also serve the purpose of verifying that CF metadata is correctly designed (if it can be correctly interpreted).*
- Extend udunits to support non-real-world calendars and formatted time. Make alternative time en/decoding software available.
- Other packages to include complete CF support e.g. CDAT, Ferret, nco.
- Implementation of CF in data formats other than netCDF e.g. xml.

Conclusions

CF is becoming widely adopted. This should help the community.

But it requires significant investment of resources (perhaps two full-time posts) in developing the CF standard and the standard name table so that they can keep up with the diversity of requirements.

Engagement is also necessary from all those with relevant expertise (scientific and technical) in providing suggestions and guidance.